

## REVIEW ARTICLE

# Machine learning and chronic kidney disease risk prediction

Marina Wainstein<sup>1</sup>, Sally Shrapnel<sup>2</sup>, Chevon Clark<sup>3</sup>, Wendy Hoy<sup>4,5</sup>, Helen Healy<sup>4,5</sup>, Ivor Katz<sup>6</sup>

<sup>1</sup>University of Queensland, Faculty of Medicine, Brisbane, Australia. <sup>2</sup>School of Mathematics and Physics, University of Queensland, Brisbane, Australia. <sup>3</sup>National Renal Care, South Africa. <sup>4</sup>CKD.QLD, Centre for Chronic Disease, University of Queensland, Brisbane, Australia. <sup>5</sup>Kidney Health Service, Royal Brisbane and Women's Hospital, Brisbane, Australia. <sup>6</sup>St George Hospital, Renal Department, Kogarah, Faculty of Medicine, University of New South Wales, Sydney, Australia.

## ABSTRACT

With a prevalence of approximately 10–15% in Africa and a close relationship with other non-communicable diseases, chronic kidney disease (CKD) can result in a significant comorbidity burden and impact on quality of life. The complex spectrum of precipitants and drivers of progression present a challenge for early diagnosis and effective interventions. Predicting this progression can provide clinicians with guidance on the need and frequency of monitoring in specialist clinics, the degree to which interventions such as kidney biopsies and aggressive risk factor modification may be of use, and to plan, in a timely manner, the various elements of dialysis initiation and transplantation. For patients, such predictions have the potential to contextualise the recommended therapies and monitoring regimes prescribed, allowing them to engage better with decision making and planning if, and when, kidney replacement therapies are needed. This paper explores the use of machine learning to facilitate such predictions and improve our understanding of CKD as well as to provide a platform for future studies to examine their clinical utility and value to both clinicians and patients.

**Keywords:** machine learning; chronic kidney disease; predictions.

## INTRODUCTION

Although infectious diseases are still the leading cause of death in Africa, non-communicable chronic diseases contribute a significant burden of morbidity and mortality [1]. The overall incidence of chronic kidney disease (CKD) in Africa is approximately 10–15%, with roughly 4.6% being advanced stages 3–5 [2]. In South Africa in 2017, there were 10,744 people in a kidney replacement therapy (KRT) programme, which represented a decline in per million population (pmp) from 70 in 1994 to 66 pmp in 2017 [3]. Nephrologists and most kidney services in Africa spend much of their time and resources triaging and managing patients with CKD and planning the delivery of care for its terminal state end-stage kidney failure (ESKF). Patients may live their entire lives unaware of having this silent disease or they may be increasingly

debilitated by its relentless progression and ultimate fate, connected to a dialysis chair or dying without it. The ability to predict such divergent outcomes is as difficult and time-consuming for most clinicians as is understanding the complexity of this disease and conceptualising the range of risk factors that contribute to its progression.

The prediction of disease trajectory and collateral events, especially when contributed by a spectrum of modifiable risk factors as does CKD, is enormously helpful and beneficial. The capacity to distinguish between patients who are likely to progress to kidney failure from those in whom the disease will linger without much consequence, could allow us to selectively deliver specialised nephrology care without overwhelming our resources, steer the urgency and aggressiveness with which we target risk fac-

tor modification and plan for dialysis and transplantation in a timely manner. On a population level, these predictions have the capacity to improve the implementation and detection rate of screening programmes and shape health policy to address at-risk populations. In research, prediction models can be used to guide entry into clinical trials and inform sample size calculations, resulting in studies with better and more pragmatic design. Ultimately, these predictions empower patients with the knowledge to plan and engage in decision making with regard to their future.

Although various risk prediction equations and scores exist today, their capacity to be adapted to wide and noisy data sets, such as electronic health records (EHR), or to model complex, non-linear relationships to predict CKD progression, is limited. Machine learning prediction models are rapidly proving their worth in many areas of medicine, from diagnosing skin lesions to directing drug trials for cancer patients [4]. These computer models, which can learn from data with minimal external programming and be deployed to find patterns in vast and complex data sets, present a unique opportunity to predict CKD trajectory and offer insights into its contributing factors. The aim of this review is, first, to look at existing algorithms that assess decline in CKD before focusing on machine learning models and their potential to predict CKD progression.

## CKD CLASSIFICATION AND THE CENTRAL ROLE OF ESTIMATED GLOMERULAR FILTRATION RATE (EGFR) AND ALBUMINURIA

The recognition of CKD as a major health priority has generated enormous efforts to improve its detection and timely management. Inherent to this task has been the development of a language to describe and characterise it according to disease severity and to provide a framework for epidemiological CKD research. In 2002, the Kidney Disease Outcomes Quality Initiative (KDOQI) of the National Kidney Foundation (NKF) published a set of guidelines aimed at defining, staging and risk stratifying CKD. The threshold eGFR of less than 60 mL/min/1.73 m<sup>2</sup> was chosen as it represented loss of half or more of the normal measured GFR in a young adult (120–130 mL/min/1.73 m<sup>2</sup>) [5], and the point at which the complications of kidney disease became apparent and significant. Five stages of CKD severity were developed, and proteinuria was identified as an important marker of kidney damage, but not formally incorporated into this earlier classification.

In 2009, the Kidney Disease: Improving Global Outcomes (KDIGO) initiated a collaborative meta-analysis to further investigate the role of low eGFR and albuminuria on mortality and kidney outcomes to inform practice guide-

lines [6]. The meta-analysis confirmed an increased risk in all outcomes (all-cause mortality, cardiovascular mortality and kidney outcomes, which included kidney failure, acute kidney injury and progressive CKD) at an eGFR less than 60 mL/min/1.73 m<sup>2</sup> as well as with a urine to albumin creatinine ratio (ACR) greater than 3 mg/mmol, independent of eGFR [6,7]. Additionally, the working group found a steep rise in risk of all outcomes below an eGFR threshold of 45 mL/min/1.73 m<sup>2</sup>, prompting a split in the stage 3 category. It was noted that risk in kidney-specific outcomes, particularly risk of progression to ESKF, were exponentially increased in patients with lower eGFR and higher albuminuria levels [6]. The revisions were implemented into the latest KDIGO CKD guidelines to generate the familiar CKD heat map that is commonly used today and cemented eGFR and albuminuria as the guiding markers of overall prognosis [8] (Figure 1, from KDIGO guidelines). They now feature heavily in most risk prediction scores for CKD progression.

Despite the value of formalising eGFR as a reliable estimate of kidney function, its routine reporting in a patient's biochemical profile generated a form of universal screening which did not achieve the desired outcome. It triggered a sharp rise in referrals to specialist kidney services, leading to longer waiting times, misdiagnosis and over-investigation of low-risk patients [9,10], especially those with extremes in muscle mass and creatinine levels. This inaccuracy generated by using eGFR as the sole marker of kidney function, added to a growing understanding of CKD as a complex and heterogeneous disease. As a result, clinicians and scientists have started to look beyond eGFR to consider other potential predictors of disease progression and to explore their integration into clinically useful prediction models.

## PREDICTING CKD PROGRESSION

Clinicians have been slow to adopt risk scores in day-to-day practice, contrasting with their increasing appearance in the clinical literature over the past 20 years. Three systematic reviews have evaluated 112 risk prediction models and scores from 1980 to 2018 [11–13].

Risk prediction models have been developed in different incident populations, from general CKD to disease-specific, such as the IgA nephropathy population [14], to stage-specific, as in patients with eGFR less than 30 mL/min/1.73 m<sup>2</sup> [15]. The risk predictions have been made for occurrence of kidney disease, for progression (either defined as a change in eGFR or reaching ESKF), as well as for the risk of cardiovascular events, hospitalisations, acute kidney injury, and all-cause mortality (Figure 2) [13].

Prognosis of CKD by GFR and albuminuria categories: KDIGO 2012				Persistent albuminuria categories Description and range		
				A1	A2	A3
				Normal to mildly increased	Moderately increased	Severely increased
				< 30 mg/g < 3 mg/mmol	30–300 mg/g 3–30 mg/mmol	> 300 mg/g > 30 mg/mmol
GFR categories (ml/min/1.73 m <sup>2</sup> ) Description and range	G1	Normal or high	≥ 90			
	G2	Mildly decreased	60–89			
	G3a	Mildly to moderately decreased	45–59			
	G3b	Moderately to severely decreased	30–44			
	G4	Severely decreased	15–29			
	G5	Kidney failure	< 15			

**Figure 1.** Prognosis of CKD by GFR and albuminuria category [6].

Green: low risk (if no other markers of kidney disease, no CKD); yellow: moderately increased risk; orange: high risk; red: very high risk. Reproduced with permission.

	Age	Sex	Ethnicity	Blood pressure	Other	Diabetes	Hypertension	Other	Histology (biopsy)	eGFR	Serum creatinine	Proteinuria / albuminuria	Serum albumin	Phosphate	Calcium	Hemoglobin	Other
General CKD	Patient characteristics					Comorbidities				Laboratory variables							
Cheng, 2017	×				X <sub>4</sub>	×	×	×					×				X <sub>2</sub>
Schroeder, 2017	×	×		×	×	×		×		×		×				×	
Hsu, 2016	×	×	×	×	X <sub>4</sub>	×		×		×		×					
Tangri, 2016 AJKD	×	×								×		×	×	×	×		×
Xie, 2016	×	×	×							×		×					
Marks, 2015	×	×			×							×					
Maziarz, 2015	×	×	×							×		×					
Levin, 2014	×	×								×		×		×		×	X <sub>3</sub>
Maziarz, 2014	×	×	×							×		×					
Drawz, 2013	×			×				×		×			×				×
Smith, 2013								×		×		×					X <sub>2</sub>
Tangri, 2011 4v model	×	×								×		×					
Tangri, 2011 8v model	×	×								×		×	×	×	×		×
Landray, 2010		×									×	×		×			
Johnson, 2008	×	×				×	×	×		×							
Johnson, 2007	×	×				×	×	×		×							
Dimitrov, 2003											×	×		×	×		

**Figure 2.** Studies and variables used in developing risk prediction scores for chronic kidney disease progression [13].

Despite this heterogeneity in model composition, many important insights can be gleaned. These included that serum creatinine can be used in place of eGFR as a marker of kidney function, that several biochemical markers (serum albumin, haemoglobin and C-reactive protein) should be examined as potential variables, that patient-specific comorbidities like diabetes and hypertension are important variables to consider either as primary kidney disease or as additional comorbidities, and that progression of CKD can be defined in multiple different ways.

## THE KIDNEY FAILURE RISK EQUATION SCORE

In 2011, Tangri and colleagues published their 4- and 8-variable risk prediction models for ESKF, jointly known as the Kidney Failure Risk Equation (KFRE), and these are generally accepted as the best prediction models for progression of kidney disease to date [16,17]. The models were developed in a development cohort and validated in a second cohort of people with eGFR less than 60 mL/min/1.73 m<sup>2</sup> (i.e., stage 3 and below). Multiple candidate variables at baseline were considered in the development of the models including age, gender, weight, systolic and diastolic blood pressure, comorbidities (cardiovascular disease and diabetes) as well as laboratory variables such as eGFR, urine ACR and serum albumin, phosphate, bicarbonate and calcium. Kidney failure, as defined by dialysis initiation or transplantation, was codified as an independent variable. Cohort participants were followed for an extended period of seven years, enabling the prospective development of 1-, 3- and 5-year time horizons for risk predictions. The best-performing model in both cohorts included age, gender, eGFR, urine ACR, serum albumin, phosphate, and bicarbonate, with a C-statistic of 0.917 (95% confidence interval 0.901–0.933) in the development cohort and 0.841 (95% CI 0.825–0.857) in the validation cohort. The models identified that lower eGFR, higher urine ACR, younger age and male gender predicted faster progression to ESKF, like the RENAAL and Kaiser Permanente models [18,19]. The 8-variable KFRE risk prediction model for ESKF added lower serum albumin, calcium and bicarbonate and higher phosphate as further predictors of progression. The lack of model performance improvement using variables of diabetic status, weight and hypertension was thought to relate to their high prevalence among the CKD population and therefore limited use as markers of disease severity.

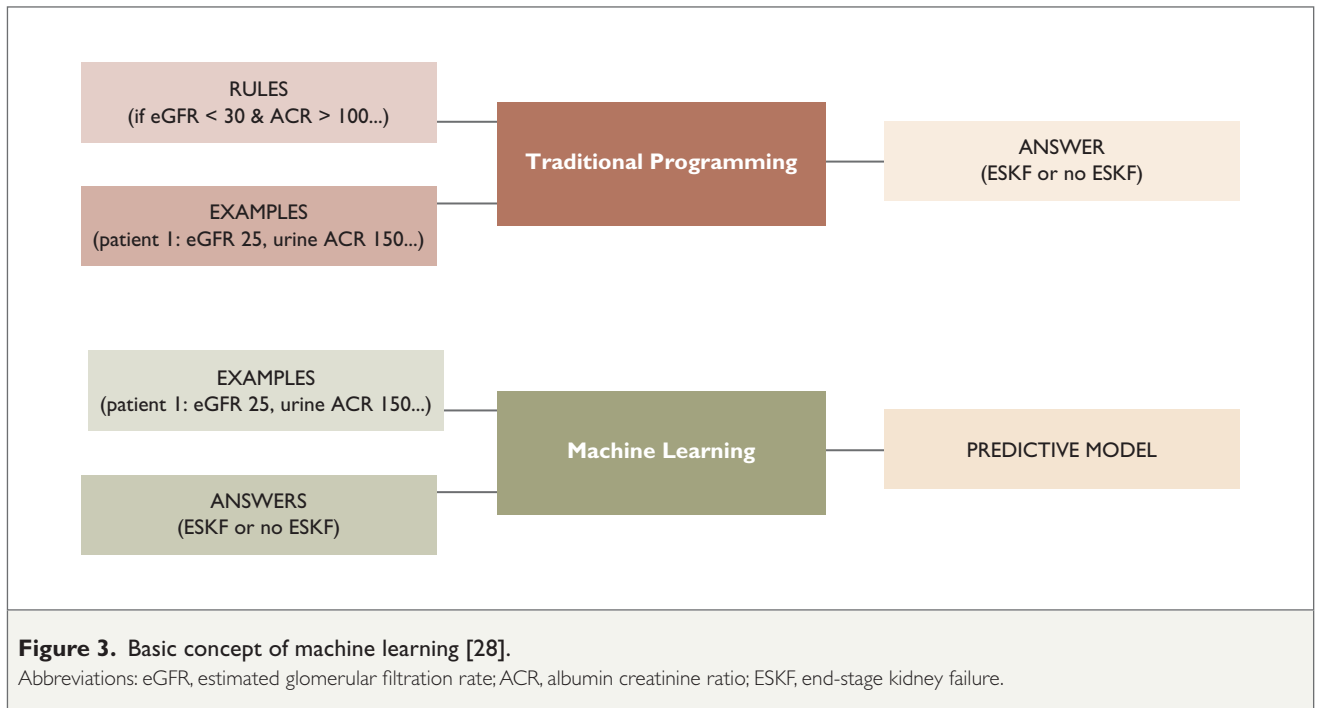
Building on this, the strength of the KFRE today lies in its extensive external validation and proposed clinical utility [20]. It is the only prediction model with proposed actionable risk thresholds to guide decision making in clinical practice, including triaging of new referrals to

specialist nephrology care and timing of pre-dialysis education and vascular access creation [21,22]. The KFRE 5-year 3% threshold was used in a population from Manitoba in 2013 to triage referrals as a response to the overwhelming number of new referrals generated by the automatic reporting of eGFR, resulting in a reduction in the waiting time to see a nephrologist from a median 280 days to 58 days [12]. This combination of good predictive performance, extensive external validation, and proven clinical utility is what makes the KFRE the current benchmark risk prediction equation in CKD.

However, several limitations and shortcomings have been specifically attributed to CKD risk prediction scores in systematic reviews. These include an inappropriate use of a heterogeneous CKD population for model development (in particular with regard to CKD aetiology), lack of specification as to when the models should be used, poor definition of prediction time frames, lack of uniformity in the outcome definition between models, and a paucity of external validation across different populations as well as longer-term impact studies, preferably in the form of a randomised, controlled trial (RCT) to confirm clinical utility [11,23,24]. The transition to an era of Big Data and precision medicine poses new challenges and demands on the development and application of new risk prediction tools, which include the computational capacity to process and integrate a large number of real-time, longitudinal predictor variables, the statistical flexibility to accommodate non-linear relationships between variables and outcomes, and the ability to be incorporated into existing patient data platforms, such as EHRs [25].

## MACHINE LEARNING

Under the broad umbrella of artificial intelligence – a science dedicated to creating intelligent computers able to perceive vision, language, and sound – emerges the field of machine learning (ML), which explores the ability of machines and computer algorithms to learn from data to make accurate and generalisable predictions. The field has evolved significantly since Arthur Samuel, a pioneer in the area of computer gaming, first coined the term in 1959 while proving that a machine could learn to play a game of checkers better and quicker than a human being [27]. At its most basic level, ML is the study of algorithms that allow computers to learn without being explicitly programmed to do so. A computer, programmed with a basic algorithmic architecture, can take a quantity of information or data (usually very large and noisy) and find generalisable predictive patterns. Importantly, this basic algorithmic structure that primes the mathematical functions a model can



learn is not constrained by any assumptions between variables and outcome in the way that traditional statistical models have been. Until recently, the use of computers to help us make predictions relied on humans inputting a set of rules, usually based on study findings, such as “if eGFR lower than 60 mL/min/1.73 m<sup>2</sup> and urine ACR greater than 30 mg/mmol, then risk of progression to ESKF equals X”, and then showing the computer new real-life patient data and asking it to produce a risk score based on the programmed rules. With ML, a computer is shown the latter patient data with the respective outcomes – that is, if and/or when they reached ESKF – to find patterns that will form the architecture of a prediction model (Figure 3).

This, coupled with the computational power to process enormous amounts of data, allows these models to consider and collate vast sources of health information to make accurate and individualised forecasts.

Broadly, ML can be divided into supervised and unsupervised learning. Supervised learning involves “training” an algorithm by exposing it to labelled data, so it can find patterns between the features of the data (independent variables) and the label – that is, dependent variable – attached to each data point, so that when it encounters those features it can correctly assign a label. Supervised ML algorithms can be divided into classification and regression types according to the nature of the outcome they try to predict – that is, whether the data belong to a certain category or they predict a numerical value. However, many of the most used ML algorithms, such as neural networks and random forest

models, can be used both for classification and regression tasks, while others such as linear regression and Naïve Bayes classifier are best suited to predict either a continuous or discrete outcome, respectively.

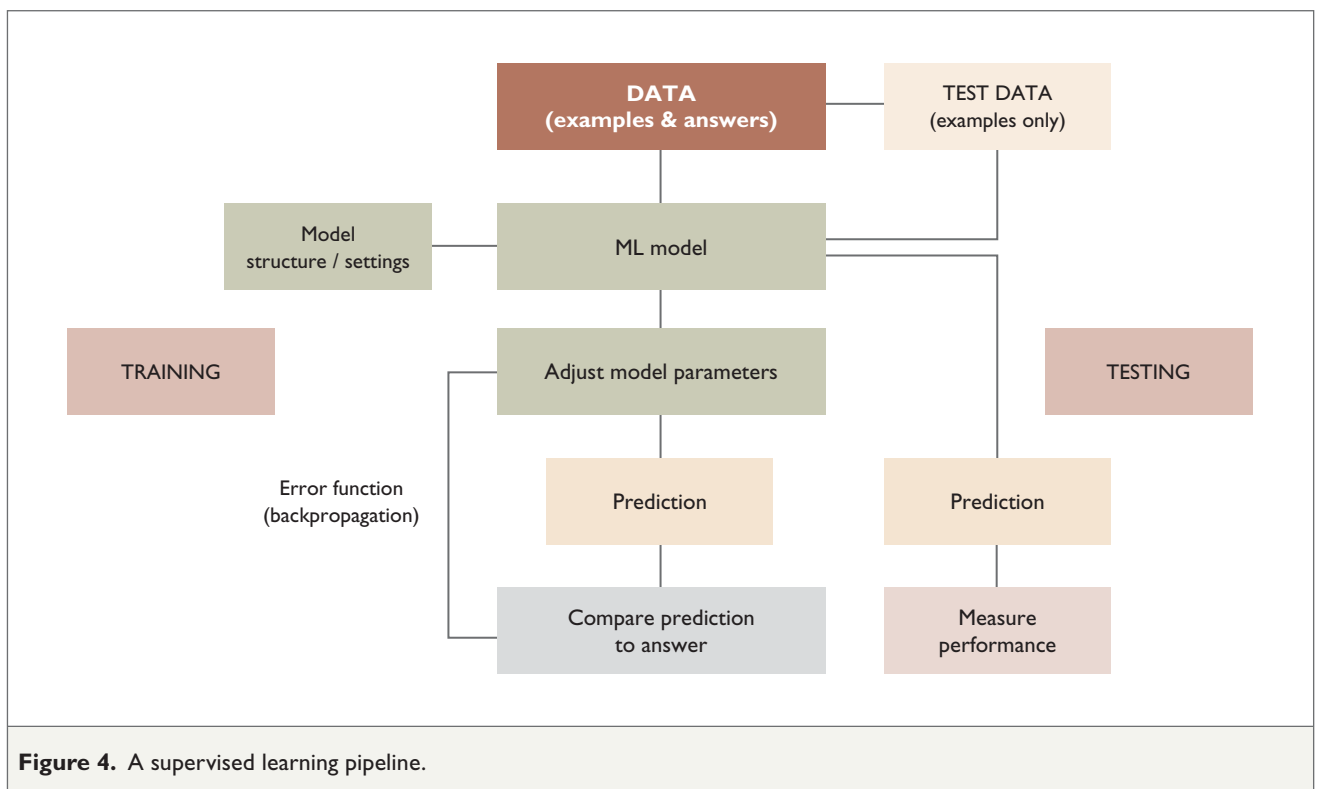
Unsupervised learning is used to find patterns in an unlabelled data set, as is the case with data mining where an ML algorithm is deployed to find any meaningful patterns or groupings in a random data set without any direction rules. An example of unsupervised learning would involve an analysis of patients’ clinical letters to find new predictors of kidney disease progression. Such a study was in fact conducted in 2016 by Singh et al., and identified ascorbic acid level and fast-food consumption as additional predictors of CKD progression [29].

Supervised learning requires training a model with a range of inputs (or features) which are associated with an outcome (or label) [30]. An example of this in nephrology would be training a model to relate a patient’s comorbidity profile – for example, the presence of hypertension, diabetes, cardiovascular disease, etc. – with the existence of kidney disease as defined by low eGFR. Once the model has been trained with enough labelled examples, it can make predictions on new and unseen data. The final trained model can be thought of as a single mathematical function that maps each input (for example, eGFR, urine ACR, age, etc.) to an outcome (for instance, ESKF or no ESKF). In this review, we will be looking at artificial neural networks (ANN) and random forest (RF) models as two examples of supervised learning.

At the start of training, the model parameters are randomised, and training follows a path of iterative improvement in reducing the error between prediction and outcome by using an optimisation technique. In the case of ANNs this is done by allowing information to be passed through layers, composed of modules or “neurons”, which change or “transform” this information according to a set of tuneable parameters (known as weights and biases) and a linear or non-linear activation function which determines if and how the information will be fed on to the next layer. After passing through several hidden layers, information reaches the final output layer and generates a prediction which is measured against the actual label using an error function. This error in prediction is then fed back through the layers (backpropagation) to allow the weights and biases to be adjusted accordingly. This process will continue iteratively and at a certain controlled rate until the prediction error is minimised [31]. In this way, a successfully trained model will have seen enough combinations of features and labels and adjusted its internal parameters to match those labels, so that when exposed to a new combination of features during the testing phase, it will correctly assign or predict the label. This pipeline of model development involving iterative training and testing forms the general process of generic supervised machine learning model development (Figure 4).

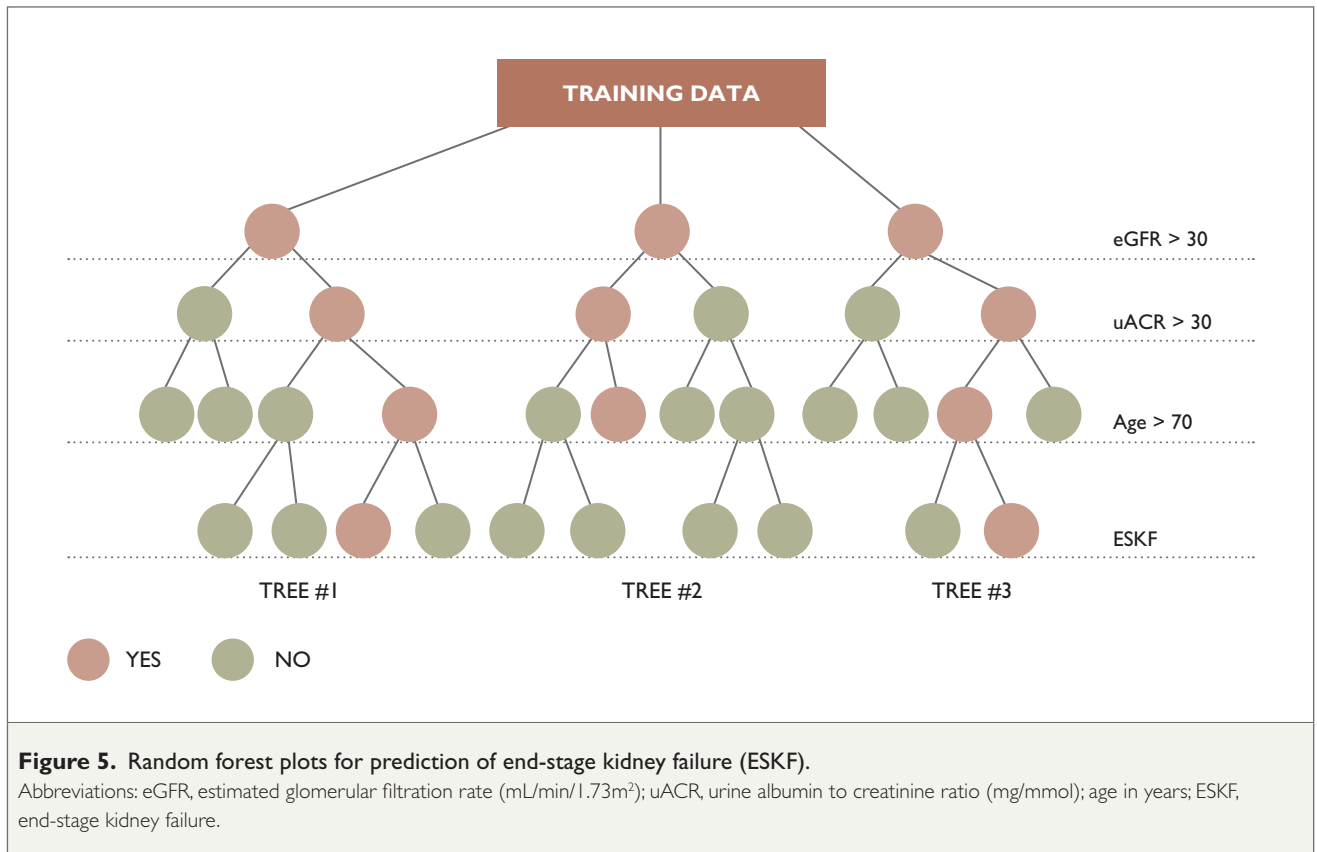
In addition, a few settings or “hyperparameters” in the model architecture, such as the number of hidden layers, number of neurons in the case of ANNs, learning rate, and optimisation function, can be adjusted to improve model performance. The architecture of stacked transformations with limitless hidden layers which act to remove any sort of parametric or frequency distribution assumptions, is what allows neural networks to model complex and powerful relationships among many variables. However, it is also what makes them less interpretable and more prone to a “black box” effect (to be discussed later).

An RF model can be used as a classification or regression algorithm and is derived from an ensemble of individual decision trees and their predictions [32,33]. In a decision tree, which is graphically represented by an upside-down tree, observations are passed down from the root through various nodes, which split the observations according to questions and subsequent decisions until a terminal node or leaf is reached (Figure 5). The questions that determine each split are drawn from the available variables in the data and selected according to their ability to split the data homogeneously, a characteristic referred to as “goodness of split” and determined through a variable’s impurity index. The goal is for each case (or patient example) to travel down the tree and be directed to a classification category (or terminal node) based on its constitutive features. In an RF model, an ensemble of trees is developed from subsets



**Figure 4.** A supervised learning pipeline.



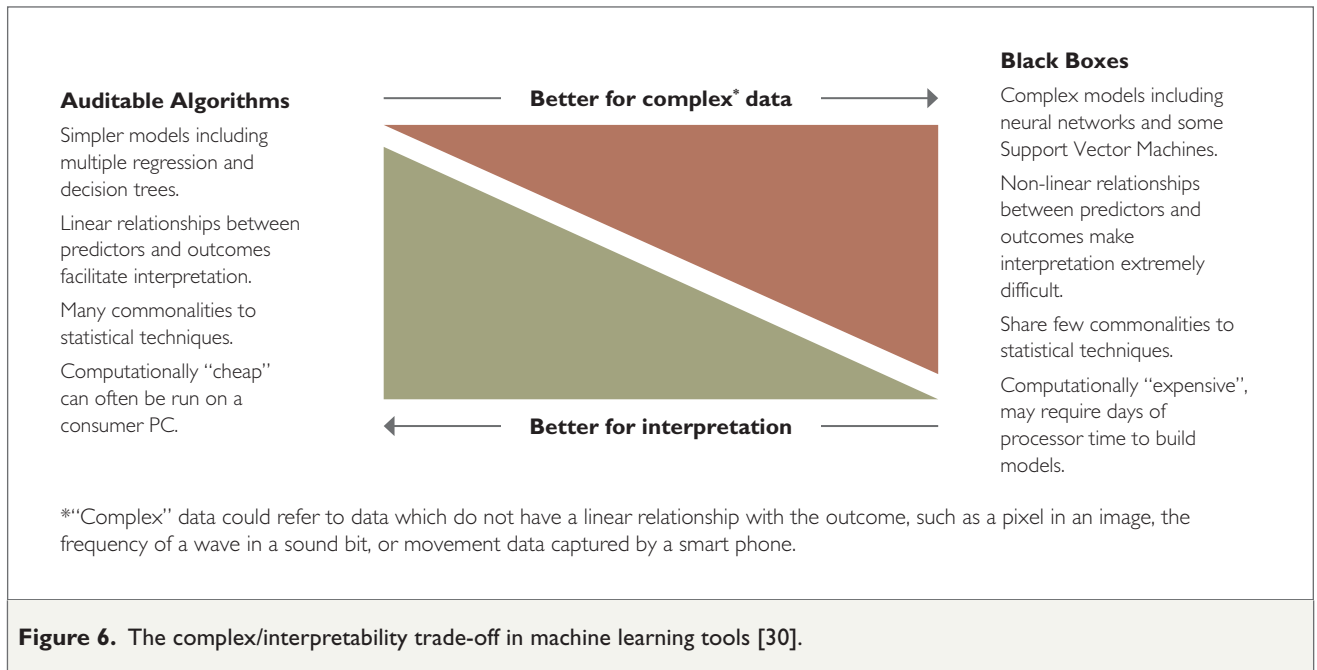


of the data which make simultaneous, individual predictions on each new case, which are then collated to form a final prediction “vote” [32]. This ensemble structure confers greater stability and generalisability to the prediction than that made by any individual tree. Performance of RFs can be improved through bagging or boosting, which allows individual trees to be developed on a randomly selected subset of the data (bagging) or iteratively using the entire training data set and adjusting weighting of samples according to classification errors (boosting) [34]. These techniques have been shown to improve prediction accuracy, lessen the probability of model overfitting, and adapt better to smaller data sets by avoiding the need to split them into training and test sets [35]. Following training, the performance of ML models can be assessed using well-known metrics of calibration, discrimination, and reclassification [34]. Additionally, RFs have the capacity to select and rank variables according to their impact on the target class, conferring additional insight and understanding of the role of individual variables on the model’s prediction.

Although we have already outlined several ways in which ML models differ from traditional statistical algorithms, on a conceptual level they can also be understood to have very different purposes. Statistical models are generally used to make inferences on relationships within a data

sample and to give a quantifiable measure of confidence that the observed relationship describes a “true” phenomenon that is not the product of noise or chance [36]. On the other hand, ML models aim to predict a future event accurately without requiring an understanding of the mechanism that links the predictive variables to the outcome. It can be said that statistics looks at the “how” and ML at the “what”. A natural conclusion stemming from this is that ML models are much less interpretable, or “auditable”, than the traditional regression and classification approaches. As an example, using an ML model to predict likelihood of requiring dialysis in a 5-year period, we may be able to predict this with great certainty and accuracy from a plethora of collected laboratory and demographic variables, but we may not be able to explain which variable/s exerted the most impact on this prediction. Having said this, ML algorithms differ in their degree of interpretability, which is largely determined by their mathematical architecture, with ANNs being the most common example of a “black box” algorithm and RF models affording greater insights and transparency (Figure 6, from Sidey-Gibbons [30]).

Choosing the right algorithm that fits a particular task is therefore of great importance. Nevertheless, interpretability and explainability of various ML models is a rapidly growing



field of research. Newer advances suggest that the trade-offs depicted in Figure 6 are somewhat misleading: techniques exist to interpret multiple aspects of ANN models, and sophisticated algorithms to analyse large complex ensembles of decision trees [37]. Such recent improvements in ML model explainability (XAI) hope to address key open problems with ML deployment. The aim is to ensure that models are fair, transparent, reliable and remain robust to data shift. While such goals also exist for traditional statistical techniques, the enormous numbers of parameters involved in deep learning and ensemble ML techniques have made XAI a fertile area of research. While it is beyond the scope of this review to provide a complete taxonomy, specific techniques exist to target training bias, treatment bias, ascertainment bias, missing data and data shift over time [38].

## ML AND RISK PREDICTION IN CHRONIC KIDNEY DISEASE

The prediction of CKD progression with ML models has gained increasing attention recently, both for the reasons already mentioned as well as because CKD is defined by a laboratory variable (eGFR) and can therefore be easily identified in EHRs. Three studies are especially relevant in the context of this review. In 2014 Rucci et al. used a classification tree analysis (CTA) to look for relevant variables associated with differential decline in eGFR [39]. The CTA used only 6 of the 17 potential predictor variables and found proteinuria to be the most discriminative variable to split the group (patients with proteinuria had a mean annual eGFR decline of  $-2.35$  vs.  $-0.80$  mL/min/1.73 m<sup>2</sup>

in patients without proteinuria). Among the group with proteinuria, those with an eGFR at baseline greater than 33 mL/min/1.73 m<sup>2</sup> appeared to have faster progression compared to an eGFR less than 33 mL/min/1.73 m<sup>2</sup> at baseline ( $-3.77$  vs  $-1.78$  mL/min/1.73 m<sup>2</sup>). The authors concluded that proteinuria was a clear predictor of CKD progression, as was phosphate to a lesser but significant extent. They further commented that eGFR alone was not sufficient to predict risk, that CKD progression slowed down with increasing age, and that diabetes was a risk factor for progression even in the absence of proteinuria.

## COMMON PROBLEMS AND LIMITATION OF MACHINE LEARNING

A common problem arises when ML algorithms are so tuned in to the particular combinations of features and outcomes that they cannot be generalised, limiting utility when presented with new data. Two ways of overcoming this problem, also known as “overfitting”, are by penalising model complexity (for example, using dropout, batch normalisation, and specific hyperparameter choices) and external validation [40]. Small data sets represent a challenge to the ML approach, particularly when the data set is split into a training and validation set, further reducing the data sets and impacting on the model's performance and generalisability.

Despite the vast potential of ML models in the modern era of Big Data to generate individualised, real-time health predictions from routinely collected data, few, if any, have been incorporated into clinical practice, particularly in the



field of nephrology. Many of the explanations for this have already been mentioned as proposed barriers to the adoption of risk scores; however, there may be other considerations such as the current state of our healthcare data platforms, the fragmentation of data capture among public and private domains, and a range of ethical considerations surrounding patient privacy and data ownership, to name a few.

The application of ML in Africa, particularly in health care and nephrology, is inhibited by legacy systems and scarce and fragmented data that are often insufficient to train ML models that can achieve good performance. This paucity of data means that models applied in Africa are often developed on non-African populations, raising the potential for unintended algorithmic bias [4]. Furthermore, limited resources, high associated costs and the poor adoption of EHRs, along with a range of ethical considerations surrounding patient privacy and data ownership, continue to contribute to a low level of digitalisation across Africa and limit the ability to integrate AI and ML technology [4]. Although computational capacity can be sourced via cloud platforms, reliable and affordable internet connectivity and electricity can be a rate-limiting step that hinders data generation and analysis needed for advanced automation of patient care. Going forward, building on existing systems rather than starting anew may facilitate overcoming many of the existing barriers in AI adoption and implementation in low-resource settings [4].

## Disclosures

None

## REFERENCES

- Matsha TE, Erasmus RT. Chronic kidney disease in sub-Saharan Africa. *Lancet Glob Health*. 2019; 7:e1587-e1588.
- George JA, Brandenburg JT, Fabian J, Crowther NJ, Agongo G, Alberts M, et al. Kidney damage and associated risk factors in rural and urban sub-Saharan Africa (AWI-Gen): a cross-sectional population study. *Lancet Glob Health*. 2019; 7:e1632-e1643.
- Davids MR, Jardine T, Marais N, Jacobs JC, Sebastian S. South African Renal Registry Report 2017. *Afr J Nephrol*. 2019; 22:60-71.
- Owoyemi A, Owoyemi J, Osiyemi A, Boyd A. Artificial Intelligence for Healthcare in Africa. *Front Digit Health*. 2020; 2.
- Davies DF, Shock NW. Age changes in glomerular filtration rate, effective renal plasma flow, and tubular excretory capacity in adult males. *J Clin Invest*. 1950; 29:496-507.
- Astor BC, Matsushita K, Gansevoort RT, van der Velde M, Woodward M, Levey AS, et al. Lower estimated glomerular filtration rate and higher albuminuria are associated with mortality and end-stage renal disease. A collaborative meta-analysis of kidney disease population cohorts. *Kidney Int*. 2011; 79:1331-1340.
- Matsushita K, Van der Velde M, Astor B, Woodward M, Levey A, De Jong P, et al. Chronic Kidney Disease Prognosis Consortium: Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: A collaborative meta-analysis. *Lancet*. 2010; 375:2073-2081.
- KDIGO group. KDIGO 2012 Clinical Practice Guidelines for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int*. 2013; 3:112-119.
- Hemmelgarn BR, Zhang J, Manns BJ, James MT, Quinn RR, Ravani P, et al. Nephrology visits and health care resource use before and after reporting estimated glomerular filtration rate. *JAMA*. 2010; 303:1151-1158.
- Glasscock RJ, Winearls C. Screening for CKD with eGFR: doubts and dangers. *Clin J Am Soc Nephrol*. 2008; 3:1563-1568.
- Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS Med*. 2012; 9:e1001344.
- Hingwala J, Wojciechowski P, Hiebert B, Buetti J, Rigatto C, Komenda P, et al. Risk-Based Triage for Nephrology Referrals Using the Kidney Failure Risk Equation. *Can J Kidney Health Dis*. 2017; 4:1-9.
- Ramspek CL, de Jong Y, Dekker FW, van Diepen M. Towards the best kidney failure prediction tool: a systematic review and selection aid. *Nephrol Dial Transplant*. 2020; 35:1527-1538.
- Barbour SJ, Coppo R, Zhang H, Liu ZH, Suzuki Y, Matsuzaki K, et al. Evaluating a New International Risk-Prediction Tool in IgA Nephropathy. *JAMA Intern Med*. 2019; 179:942-952.
- Grams ME, Sang Y, Ballew SH, Carrero JJ, Djurdjev O, Heerspink HJL, et al. Predicting timing of clinical outcomes in patients with chronic kidney disease and severely decreased glomerular filtration rate. *Kidney Int*. 2018; 93:1442-1451.
- Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*. 2011; 305:1553-1559.
- Whitlock RH, Chartier M, Komenda P, Hingwala J, Rigatto C, Walld R, et al. Validation of the Kidney Failure Risk Equation in Manitoba. *Can J Kidney Health Dis*. 2017; 4:2054358117705372.
- Keane WF, Zhang Z, Lyle PA, Cooper ME, de Zeeuw D, Grunfeld JP, et al. Risk scores for predicting outcomes in patients with type 2 diabetes and nephropathy: the RENAAL study. *Clin J Am Soc Nephrol*. 2006; 1:761-767.
- Johnson ES, Thorp ML, Platt RW, Smith DH. Predicting the risk of dialysis and transplant among patients with CKD: a retrospective cohort study. *Am J Kidney Dis*. 2008; 52:653-660.
- Tangri N, Grams ME, Levey AS, Coresh J, Appel LJ, Astor BC, et al. Multinational Assessment of Accuracy of Equations for Predicting Risk of Kidney Failure: A Meta-analysis. *JAMA*. 2016; 315:164-174.
- Hingwala J, Bhangoo S, Hiebert B, Sood MM, Rigatto C, Tangri N, et al. Evaluating the implementation strategy for estimated glomerular filtration rate reporting in Manitoba: the effect on referral numbers, wait times, and appropriateness of consults. *Can J Kidney Health Dis*. 2014; 1:9.
- Tangri N, Ferguson T, Komenda P. Pro: Risk scores for chronic kidney disease progression are robust, powerful and ready for implementation. *Nephrol Dial Transplant*. 2017; 32:748-751.
- Ramspek CL, de Jong Y, Dekker FW, van Diepen M. Towards the best kidney failure prediction tool: a systematic review and selection aid. *Nephrology, dialysis, transplantation* 2020; 35:1527-1538.
- Dekker FW, Ramspek CL, van Diepen M. Con: Most clinical risk scores are useless. *Nephrol Dial Transplant*. 2017; 32:752-755.
- Saez-Rodriguez J, Rinschen MM, Floege J, Kramann R. Big science and big data in nephrology. *Kidney Int*. 2019; 95:1326-1337.
- Topol E. Deep medicine: how artificial intelligence can make healthcare human again. ed. New York: Hachette Book Group; 2019.
- Samuel A. Some Studies in Machine Learning Using the Game of Checkers. *IBM J Res Dev*. 1959; 3:210-229.
- Datalya. Machine Learning vs. Traditional Programming Paradigm. <https://datalya.com/blog/machine-learning/machine-learning-vs-traditional-programming-paradigm>. Accessed 12/07/2021.

29. Singh K, Betensky RA, Wright A, Curhan GC, Bates DW, Waikar SS. A Concept-Wide Association Study of Clinical Notes to Discover New Predictors of Kidney Failure. *Clin J Am Soc Nephrol*. 2016; 11:2150-2158.
30. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol*. 2019; 19:64.
31. Wainberg M, Merico D, Delong A, Frey BJ. Deep learning in biomedicine. *Nat Biotechnol*. 2018; 36:829-838.
32. Breiman L. Random forests. *Machine Learning*. 2001; 45:5-32.
33. Banerjee M, Reynolds E, Andersson HB, Nallamothu BK. Tree-Based Analysis: A Practical Approach to Create Clinical Decision-Making Tools. *Circulation: Cardiovascular Quality and Outcomes*. 2019; 12:e004879.
34. Breiman L. Bagging predictors. *Machine Learning*. 1996; 24:123-140.
35. Briem GJ, Benediktsson JA, Sveinsson JR. Multiple classifiers applied to multisource remote sensing data. *IEEE Trans Geosci Remote Sens*. 2002; 40:2291-2299.
36. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018; 15:233-234.
37. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. Explainable AI for trees: From local explanations to global understanding. *arXiv preprint arXiv:190504610*. 2019.
38. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Benetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*. 2020; 58:82-115.
39. Rucci P, Mandreoli M, Gibertoni D, Zuccala A, Fantini MP, Lenzi J, et al. A clinical stratification tool for chronic kidney disease progression rate based on classification tree analysis. *Nephrol Dial Transplant*. 2014; 29:603-610.
40. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. ed: MIT Press; 2016.